

# Mechanism for Missing Data Incorporated into Joint Modelling of Ordinal Responses

Anna Ivanova

*I-BioStat, KU Leuven, Belgium.*

Geert Molenberghs

*I-BioStat, Universiteit Hasselt, Belgium; I-BioStat, KU Leuven, Belgium.*

Geert Verbeke

*I-BioStat, KU Leuven, Belgium; I-BioStat, Universiteit Hasselt, Belgium.*

**Summary.** We analyze the problem of two clinically inseparable, repeatedly measured responses of ordinal type by also incorporating their missingness process. **In our application these are the therapeutic effect and extent of side effects of fluvoxamine.** In case of a composite endpoint, the scientific questions addressed can be answered only when the responses are modeled jointly. As an extension of the methodology, several missing not at random models were fitted to a set of observed data and shown to approximately yield the same result as their missing at random counterparts, although it affects precision. In addition, the effect of various identifying restrictions on multiple imputation is investigated. An alternative numerical approximation method is suggested to reduce computational time.

**Some Keywords:** Fluvoxamine trial; Proportional odds mixed model; Missing not at random; Sensitivity analysis; Multiple imputation.

## 1. Introduction

When conducting longitudinal studies, researchers have to deal with the complex structure of the data as well as with the presence of incompleters. Also, some data types can turn out to be more difficult to model than others. In addition, in many studies, especially in the medical realm, several endpoints could not be regarded as separate responses but should be lumped into a single composite endpoint.

This work is motivated by the fluvoxamine trial, comprising data with two inseparable responses of ordinal type, measured repeatedly over time, and where for some subjects values are missing. When modelling, the first aspect that needs to be taken into account is the longitudinal nature of the data. Many statistical models have been developed for analyzing longitudinal data for which the random-effects approach has been very popular for several decades (Laird and Ware (1982), Breslow and Clayton (1993), Wolfinger and O'Connell (1993) and Engel and Keen (1994)). In the fluvoxamine trial, after treating a patient with the drug, to characterize a patient's condition, the therapeutic

effect and side effects were measured simultaneously. In such a context, joint modelling is preferred over a separate analysis of different responses given that the former also allows every outcome to have its own random effects and the association between different outcomes is captured by the correlation between the random effects. The strong advantage is that it enables researcher to obtain answers to various research questions, all from the same model. A wide variety of joint modelling techniques have already been described: Thiébaud *et al.* (2002), Iddi and Molenberghs (2012), Rizopoulos 2012, Fieuws and Verbeke (2006), Fieuws *et al.* (2006), Fieuws *et al.* (2008), Horrocks and van den Heuvel (2009), Vasdekis *et al.* (2014), Molenberghs and Verbeke (2005). For an excellent, relatively early review, we refer to the work of Tsiatis and Davidian (2004). However, the joint modelling of ordinal responses received little attention.

**Since in the fluvoxamine trial, some of the patients did not appear at all scheduled appointments, as is often the case with empirical research, we had to deal with incomplete data.** From a methodological point of view, several aspects are important, first of all, the modelling framework: *selection* (SeM), *pattern-mixture* (PMM), and *shared-parameter modelling* (SPM) (for an overview, see *e.g.* Molenberghs and Kenward 2007). Second, the mechanisms of missingness (Rubin (1976)): the data can be *Missing Completely at Random* (MCAR) when the missingness is independent of both the observed and unobserved data; *Missing at Random* (MAR) when, conditional on the observed data, missingness is independent of the unobserved measurements; *Missing Not at Random* (MNAR) when MCAR and MAR do not hold. In the PMM framework, the *Multiple imputation* (MI, Rubin (1978), van Buuren (2007)) technique can be applied **to analyze conveniently the data**. The main idea of MI is to investigate the effect of different identifying-restrictions strategies for the incompleters.

### 1.1. Fluvoxamine Trial

The data are from a multicentre study involving 315 patients that were treated by fluvoxamine for psychiatric symptoms stemming from a dysregulation of serotonin in the brain. The data are discussed in Molenberghs and Lesaffre (1994), Kenward *et al.* (1994), Molenberghs *et al.* (1997), Michiels and Molenberghs (1997), and Molenberghs and Verbeke (2005). After their recruitment into the study, patients were assessed at four visits. The therapeutic effect and the extent of side effects were scored at each visit on an ordinal scale. The side effect response is coded as 1: none; 2: not interfering; 3: interfering significantly with functionality; 4: side effects surpasses the therapeutic effect. Similarly, the effect of therapy is recorded on a four point ordinal scale: 1: no improvement or worsening; 2: minimal improvement; 3: moderate improvement and 4: important improvement. Hence, a side effect is present when new symptoms occur and a therapeutic effect whenever old symptoms disappear. A total of 299 patients with monotone sequences have at least one measurement, including 224 completers. A summary is given in Table 1. There is also baseline covariate information for

**Table 1.** Fluvoxamine Trial. Number of observations with side and therapeutic effect categories for each of the four time points

Ther. Effect	# Observations			
	Week 2	Week 4	Week 8	Week 12
0	19	64	110	135
1	95	114	93	62
2	102	62	30	19
3	83	29	10	10
Side Effect	Week 2	Week 4	Week 8	Week 12
0	128	144	156	148
1	128	103	79	71
2	28	17	6	7
3	15	5	2	0

each subject, including gender, age, presence of psychiatric antecedents, initial severity of the disease, duration of the actual mental illness. Obviously, one expects that the improvement of the patient's condition as measured by the therapeutic effect is correlated with the side effects of the therapy, and this association can only be studied if both outcomes are jointly modeled.

Furthermore, some of the patient profiles had one or more missing observations for the therapeutic and side effects. Note that, as these two outcomes were measured simultaneously, if a patient has a missing value for the therapeutic effect, he/she will also have one for side effects, and vice versa. The patterns of the missing data are summarized in Table 2. Among the incomplete sequences, the dropout patterns are the more common ones. There are also 2 late-entry patients. Observe that there are 14 subjects in total without any follow-up measurements. The latter are still an integral part of the trial, as they provide baseline information, including covariate information and baseline assessment of severity of mental illness. The scientific questions that are posed, considering the presence of missingness, can be formulated as follows: Can the different assumptions about the mechanism of the missingness and also the different identifying restrictions for the missing values lead to different results?

## 1.2. Joint Modelling of Ordinal Response(s) and Process of Missingness

First, we will specify a model for a single ordinal response. A special case of GLMM, which is of particular interest to this work, is the proportional odds mixed model (POMM) for ordinal outcomes (Agresti and Lang 1993).

Consider a longitudinal outcome of ordinal type. Let  $Y_{ij}$  be the  $j$ th measurement of  $i$ th subject ( $i = 1, \dots, N$ ,  $j = 1, \dots, n_i$ ) with values  $r = 1, \dots, R$ . Independence across subjects is further

**Table 2.** Fluvoxamine Trial. Missing data patterns for side and therapeutic effects where 'x' denotes an observed measurement and '?' a missing longitudinal response

Patient	Week 2	Week 4	Week 8	Week 12	# Observations
1	x	x	x	x	224
2	x	x	x	?	18
3	x	x	?	?	26
4	x	?	?	?	31
5	?	?	?	?	14
6	?	?	?	x	1
7	?	x	x	x	1

assumed. We first define  $R$  indicator variables as:

$$W_{r,ij} = \begin{cases} 1 & \text{if } Y_{ij} = r, \\ 0 & \text{otherwise.} \end{cases}$$

Evidently, these are redundant but any subset of  $R - 1$  components is not. Group the dummies into vectors  $\mathbf{W}_{ij}$  for a specific subject  $i$  and occasion  $j$ , and further into  $\mathbf{W}_i$  for all dummies across all occasions for subject  $i$ . We further assume a multinomial distribution  $\mathbf{W}_{ij} \sim \text{multinomial}(\boldsymbol{\pi}_{ij})$ , with  $\boldsymbol{\pi}_{ij} = (\pi_{1,ij}, \dots, \pi_{r,ij}, \dots, \pi_{R,ij})$ . The multinomial distribution at a given occasion is determined by the modelling choice made for the ordinal outcome. The probabilities can be written as:

$$\pi_{r,ij} = \begin{cases} \kappa_{1,ij} & \text{if } r = 1, \\ \kappa_{r,ij} - \kappa_{r-1,ij} & \text{if } 1 < r < R, \\ 1 - \kappa_{R-1,ij} & \text{if } r = R, \end{cases}$$

where, assuming proportional odds:

$$\kappa_{r,ij} = \frac{\exp(\xi_{0r} + \mathbf{x}'_{ij}\boldsymbol{\xi} + \mathbf{z}'_{ij}\mathbf{b}_i)}{1 + \exp(\xi_{0r} + \mathbf{x}'_{ij}\boldsymbol{\xi} + \mathbf{z}'_{ij}\mathbf{b}_i)}.$$

Here,  $\xi_{01} \leq \dots \leq \xi_{0(R-1)}$  are intercepts,  $\boldsymbol{\xi}$  fixed regression coefficients,  $\mathbf{b}_i$  a vector of normally distributed random effects, and  $\mathbf{x}_{ij}$  ( $\mathbf{z}_{ij}$ ) the design vector for the fixed (random) effects at occasion  $j$ . Second, to deal with missingness, the subject-specific outcomes are summarized into a vector  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})'$  with observed ( $\mathbf{Y}_i^o$ ) and missing components ( $\mathbf{Y}_i^m$ ). In addition, for each measurement  $j$ , we define the *missing data indicator* as follows:

$$R_{ij} = \begin{cases} 1 & \text{if } Y_{ij} \text{ is observed,} \\ 0 & \text{otherwise.} \end{cases}$$

Then, the  $R_{ij}$ s can be grouped into a vector of missing data indicators  $\mathbf{R}_i = (R_{i1}, R_{i2}, \dots, R_{in_i})'$ , which is evidently of the same length as  $\mathbf{Y}_i$ .

To incorporate the missingness process into the modelling of an ordinal response, a joint model is built by describing the joint density  $f(\mathbf{Y}_i, \mathbf{R}_i)$  of vectors  $\mathbf{Y}_i$  and  $\mathbf{R}_i$ . A very flexible way forward is by considering a mixed model for both outcomes and to allow the random effects in these to be correlated. This allows for joint modelling of outcomes of a different nature and does not alter the interpretation of the parameters in the sub-models for individual outcomes. This approach has been applied already by Fieuws and Verbeke (2006) and Iddi and Molenberghs (2012), but not yet in combination with ordinal outcomes.

For the case of an ordinal outcome, a POMM is assumed, as described before, while for the binary missing data indicator a GLMM is assumed. Note also that the covariates in both models may, but do not need to be the same. Let  $\mathbf{b}_{1i}$  and  $\mathbf{b}_{2i}$  be the vectors of random effects for the first and second responses, respectively. A joint model is now obtained by assuming a joint distribution for both sets of random effects  $\mathbf{b}_{1i}$  and  $\mathbf{b}_{2i}$ . More specifically, it will be assumed that  $\mathbf{b}_i = (\mathbf{b}_{1i}, \mathbf{b}_{2i})'$  is normally distributed with zero mean and covariance matrix  $G$ ; the latter contains components to model the association within each outcome sequence separately, as well as to model the association between sequences. It will also be assumed that, conditionally on  $\mathbf{b}_i$ , the vectors  $\mathbf{Y}_i$  and  $\mathbf{R}_i$  are independent, i.e., we assume that their association is completely captured by the association between the random effects.

Often a general unstructured matrix  $G$  is assumed, but specific restrictions can be imposed as well. For example, perfect correlations between some of the elements in  $\mathbf{b}_{1i}$  and those in  $\mathbf{b}_{2i}$  would lead to a joint model in which the random effects are shared between the two outcomes (Molenberghs and Verbeke 2005). Evidently, this is equivalent to sharing components between  $\mathbf{b}_{1i}$  and  $\mathbf{b}_{2i}$ .

The estimation and inferences for all parameters in this joint model are based on standard likelihood theory.

Further, an extension can be made to two or more ordinal outcomes. If all of them were measured simultaneously, such as the therapeutic effect and the side effects from the fluvoxamine trial, the missing data indicator will be common to all outcomes, e.g., in the fluvoxamine trial,  $\mathbf{R}_{1i} \equiv \mathbf{R}_{2i} \equiv \mathbf{R}_i$ .

Considering as an example Table 2 of the fluvoxamine trial, we can distinguish two patterns of missingness, *monotone* and *non-monotone*. The former corresponds to the so-called *dropouts* –the subject was excluded from the study before completion–, the latter to the *late entries* –some of the initial response measurements were missing.

When restricting missingness to dropouts, the missing indicator  $\mathbf{R}_i$  takes the form of  $(1, \dots, 1, 0, \dots, 0)'$  and therefore can be aggregated into the occasion of the last observed measurement  $D_i$ :

$$D_i = \sum_{j=1}^{n_i} R_{ij}.$$

Now, assume that we model the outcome  $\mathbf{y}_i$  and the missingness  $\mathbf{r}_i$  via two different sets of parameters:  $\boldsymbol{\theta}$  and  $\boldsymbol{\psi}$ , respectively. Also, assume a common set of covariates  $\mathbf{x}_i$ , consisting of the union of the

covariates influencing the two processes. Then,  $f(\mathbf{y}_i, \mathbf{r}_i | \mathbf{x}_i, \boldsymbol{\theta}, \boldsymbol{\psi})$  is the joint density of measurement and missingness as described before. Different factorizations of this joint density lead to different frameworks.

The *selection model* (SeM; Rubin 1976, Little and Rubin 2002) is based on the following factorization of the joint density:  $f(\mathbf{y}_i, \mathbf{r}_i | \mathbf{x}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\psi})$ .

An alternative framework is the so-called *pattern-mixture modelling* (PMM; Little 1993, 1994):  $f(\mathbf{y}_i, \mathbf{r}_i | \mathbf{x}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i | \mathbf{x}_i, \mathbf{r}_i, \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{x}_i, \boldsymbol{\psi})$ . PMM can be seen as a mixture of different populations, characterized by the given pattern of missingness.

Further, instead of using SeM and PMM frameworks, the outcome and missingness processes can be jointly modelled using a *shared-parameter model* (SPM; Wu and Carroll 1988, Follmann and Wu 1995, Little 1995, Njagi *et al.* 2014, Creemers *et al.* 2010), where the existence of a vector of shared parameters  $\mathbf{b}_i$  is assumed:  $f(\mathbf{y}_i, \mathbf{r}_i | \mathbf{x}_i, \mathbf{b}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i | \mathbf{x}_i, \mathbf{b}_i, \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{x}_i, \mathbf{b}_i, \boldsymbol{\psi})$ .

A taxonomy of missing data mechanisms was introduced by Rubin (1976) and further developed by Little and Rubin (2002). It is based on the following conditional density of the missing process  $\mathbf{R}_i$ :  $f(\mathbf{r}_i | \mathbf{x}_i, \mathbf{y}_i^o, \mathbf{y}_i^m, \boldsymbol{\psi})$  and is easily developed in the selection setting. As a result, three types of missing data mechanism have been defined (Verbeke and Molenberghs 2000, Fitzmaurice *et al.* 2004, Molenberghs and Kenward 2007): **Missing Completely at Random (MCAR)** when the missingness is independent of both observed and unobserved data, which corresponds to the factorization  $f(\mathbf{r}_i | \mathbf{x}_i, \mathbf{y}_i^o, \mathbf{y}_i^m, \boldsymbol{\psi}) = f(\mathbf{r}_i | \mathbf{x}_i, \boldsymbol{\psi})$ , **Missing at Random (MAR)** when conditional on the observed data the missingness is independent of the unobserved measurements,  $f(\mathbf{r}_i | \mathbf{x}_i, \mathbf{y}_i^o, \mathbf{y}_i^m, \boldsymbol{\psi}) = f(\mathbf{r}_i | \mathbf{x}_i, \mathbf{y}_i^o, \boldsymbol{\psi})$ , and **Missing Not at Random (MNAR)** when neither MCAR nor MAR applies. Here, the conditional density will abide to the general form:  $f(\mathbf{r}_i | \mathbf{x}_i, \mathbf{y}_i^o, \mathbf{y}_i^m, \boldsymbol{\psi})$ .

Also, it was shown that, under the MAR assumption for modelling under the separability conditions of the parameters  $\boldsymbol{\theta}$  and  $\boldsymbol{\psi}$ , the inference for these parameters based on likelihood is valid. This property is known as *ignorability*, and brings us to the so-called *direct likelihood analysis*.

Another method for incorporating missingness into the data analysis is multiple imputation (MI) that was formally introduced by Rubin (1978), who later published a comprehensive overview (Rubin 1987). For the continued development of the MI procedure we refer to sources such as Rubin and Schenker (1986), Little and Rubin (2002), Schafer (2003), and van Buuren (2007). In Verbeke and Molenberghs (2000), Molenberghs and Verbeke (2005), Molenberghs and Kenward (2007), Carpenter and Kenward (2013) where several classical and more advanced MI techniques are described.

The main idea behind MI is to replace each missing measurement by a number of imputed values, say  $M$ . These values are drawn from the distribution that incorporates the uncertainty about the value to impute. Then, each one of the  $M$  imputed data sets is analyzed separately by using standard procedures for complete data and the results of the analysis combined into a single one. In Rubin's framework (1987), the imputations are drawn from the posterior distribution of the *missing given*

*the observed values.*

As in selection modelling, a sensitivity analysis can also be conducted within the pattern-mixture family. When investigating the structure of longitudinal data, we can conclude that the non-completers stand out from the completers, especially in the case of early dropouts. Hence, it is common to borrow information from other or even from all available patterns. For example, the identifying restriction termed *complete case missing values* (CCMV), proposed by Little (1993), can be assumed. Here, the unavailable information is always borrowed from the completers. For this strategy to be stable there should be a sufficiently large number of completers. Alternatively, the nearest identified pattern can be used. This restriction is referred to as *neighboring case missing values* or NCMV (Verbeke and Molenberghs 2000, Molenberghs and Kenward 2007). However, using information only from one pattern can be insufficient, and hence, one can base the identification on all available patterns for which the previous component is identified. This restriction is called *available case missing value* (ACMV; Verbeke and Molenberghs 2000) and also corresponds to MAR, whereas CCMV and NCMV are MNAR restrictions.

In contrast to Rubin’s MI based on joint modelling of multiple longitudinal processes, *Full Conditional Specification* (FCS) factorizes the multivariate distribution into a set of conditional densities  $f(y_j|y_1, \dots, y_{j-1})$  for each variable  $Y_j$ . Strictly speaking, the conditional densities do not guarantee the existence of a joint distribution of all variables. However, if we start from the well-specified joint distribution, then the conditional ones can be straightforwardly derived from it. **The key assumption of FCS is ignorability – the conditional densities for  $Y_j^m$  given  $Y_j^o$  do not depend on missing patterns– but the method can be easily extended to MNAR when specifying the number of identifying restrictions.**

As to the choice of the number of imputations, the early literature suggests that a small number  $M$  of imputations, generally in the 3–5 range, is adequate because the larger the number of imputations the smaller the increase in efficiency (Rubin and Schenker 1986, Rubin 1987). However, recent publications pointed out that small  $M$  values lead to results that are affected by Monte Carlo error: using the same  $M$ , the point and precision estimates may vary considerably if the MI analysis is repeated with a different set of imputations (White *et al.* 2011, van Buuren 2012), especially when outcomes are non-Gaussian and/or focus is on hypothesis testing. Hence, a larger number of imputations is preferable.

## 2. Analysis of Fluvoxamine Trial

### 2.1. Direct Likelihood Analysis

Recall from Section 1.1 that the study encompasses four time points, at each one of which side effects and therapeutic effect are assessed on a 4-point ordinal scale. Note that the missing-data patterns occurring are mostly of the monotone type. For this reason we exclude the two non-monotone

sequences.

Let  $\mathbf{Y}_{kij}$  be the score for patient  $i$  at time point  $j$ ,  $k = 1, 2$  for therapeutic and side effects. The proportional odds logistic regression with random intercept for the  $k$ th response can be expressed as follows:

$$\begin{aligned} \text{logit}[P(\mathbf{Y}_{kij} \leq r)] &= \xi_{k,0r} + b_{ki} + \xi_{k,11}t_{1ij} + \xi_{k,12}t_{2ij} + \xi_{k,13}t_{3ij} + \xi_{k,2}X_{1i} \\ &\quad + \dots + \xi_{k,5}X_{4i}, \end{aligned} \quad (1)$$

where  $t_{1ij}$ ,  $t_{2ij}$ , and  $t_{3ij}$  are dummies corresponding to weeks 4, 8, and 12, respectively. Clinically, one cannot separate therapeutic effect from side effects, because the former refers to the disappearance of existing symptoms, whereas the latter corresponds to emerging symptoms. To capture the association between the responses, various assumptions about the distribution of the random effects can be made. We will model this association through the following assumptions for the random effects:

$$\begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{pmatrix} \right]. \quad (2)$$

The fit is presented as Model 0 in Tables 3 and 4. This analysis is performed under ignorability.

## 2.2. Analysis under Missing Not at Random Assumption

Next, we perform a sensitivity analysis under MNAR. Without loss of generality, we assume that the missingness process can be modeled using the same set of covariates as the two outcomes of scientific interest. The following logistic regression can be specified for the missing indicator:

$$\text{logit}[P(D_i > j)] = \xi_{3,0} + b_{3i} + \xi_{3,11}t_{1ij} + \xi_{3,12}t_{2ij} + \xi_{3,13}t_{3ij} + \xi_{3,2}X_{1i} + \dots + \xi_{3,5}X_{4i}. \quad (3)$$

Let the side and therapeutic effects be modeled as in (1). To capture the association between (1) and (3), we consider the following options:

- (1) allowing  $b_{3i}$  to be pairwise correlated with  $b_{1i}$  and  $b_{2i}$  (Model 1);
- (2) shared modelling by considering random effects for both outcomes  $b_{1i}$  and  $b_{2i}$ , shared with the missing indicator, scaled by factors  $\lambda_1$  and  $\lambda_2$ . This corresponds to Model 2 in Tables 3 and 4. Hence, the logistic regression for the missing data indicator takes the following form:

$$\begin{aligned} \text{logit}[P(D_i > j)] &= \xi_{3,0} + \lambda_1 b_{1i} + \lambda_2 b_{2i} + \xi_{3,11}t_{1ij} + \xi_{3,12}t_{2ij} + \xi_{3,13}t_{3ij} + \xi_{3,2}X_{1i} \\ &\quad + \dots + \xi_{3,5}X_{4i}; \end{aligned}$$

- (3) modelling the association between missingness and every outcome by including two extra random intercepts:  $b_{3i}$  for the therapeutic effect and  $b_{4i}$  for the side effects. Further, the missingness process will be associated with the outcomes through these shared random intercepts  $b_{3i}$  and



$b_{4i}$ , which are again included using scale factors  $\lambda_3$  and  $\lambda_4$ . Then, the model for the missingness process can be written as follows:

$$\begin{aligned} \text{logit}[P(D_i > j)] &= \xi_{3,0} + \lambda_3 b_{3i} + \lambda_4 b_{4i} + \xi_{3,11} t_{1ij} + \xi_{3,12} t_{2ij} + \xi_{3,13} t_{3ij} + \xi_{3,2} X_{1i} \\ &\quad + \dots + \xi_{3,5} X_{4i}. \end{aligned}$$

See Model 3 in Tables 3 and 4.

These models exhibit the following hierarchical relationships: the most complex model is model 3, containing model 1 as a sub-model, with model 2 a further sub-model of the latter. For more details about model-specific variance-covariance matrices of the random effects structure, see Supplementary Materials B, matrices (4)–(6). The most complex model 3 is overspecified, but can be used as a basis for sensitivity analysis, and as a resource to select identified sub-models.

Model 1 had computational issues and was further discarded. Models 2 and 3 were implemented in the NLMIXED procedure of SAS 9.4, where numerical integration adaptive Gaussian quadrature was used with  $Q = 10$ . We paid special attention to fitting Model 3 given its complexity. The results will be discussed in Section 2.3.

When comparing the fitted models of Tables 3 and 4, we observe that the parameter estimates for the fixed effects of the therapeutic and the side effects are approximately the same for all three models. As to the random part, Model 2 yields slightly larger estimates than Model 0, whereas Model 3 generates slightly lower estimates than Model 0. The similarity of the results for Model 0 under MAR and Models 2 and 3 under MNAR confirms the observation of Molenberghs *et al.* (2008). These authors showed a strong connection between the MNAR model fit and the MAR model fit. The most important conclusion was that the MNAR model fitted to a set of observed data can be reproduced by a MAR counterpart, which is in our case Model 0. However, sometimes it happens that the parametric MAR model does not properly fit the observed data. It is therefore of interest to fit a sufficiently different MNAR model to ensure a good fit to the observed data and then examine the corresponding MAR version. In addition, these authors concluded that the MNAR and MAR models produce different predictions for unobserved outcomes, given the observed ones.

Further observations concern the standard errors of the estimates: given that the missing data mechanism is incorporated, the standard error estimates of all parameters increase compared to those of the model under MAR. It means that Models 2 and 3 are less efficient than Model 0, and in case of prediction the uncertainty will be larger.

### 2.3. Additional Investigation of Model 3

Implementing Model 3 leads to some computational problems due to its complexity. Here, the therapeutic and side effects are modeled jointly as described in Section 2.1 and, to add the missingness process to this joint modelling, two extra random intercepts:  $b_{3i}$  for the therapeutic effect, and  $b_{4i}$  for

**Table 3.** Fluvoxamine Trial. Response-specific parameter estimates (s.e.) of the regression coefficients. Model 0: model with two outcomes only; Model 2: model with two random intercepts both of which are shared with missing data indicator; Model 3: model with four random intercepts of which two are shared with missing data indicator

		Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Effect	Par.	Model 0	Model 2	Model 3
Therapeutic Effect				
int. 0	$\xi_{1,00}$	−2.16 (1.18)	−2.25 (1.23)	−2.27 (1.23)
int. 1	$\xi_{1,01}$	1.03 (1.18)	0.93 (1.26)	0.91 (1.26)
int. 2	$\xi_{1,02}$	3.42 (1.19)	3.33 (1.31)	3.31 (1.32)
time (week = 4)	$\xi_{1,11}$	2.07 (0.20)	2.04 (0.20)	2.04 (0.20)
time (week = 8)	$\xi_{1,12}$	3.60 (0.25)	3.55 (0.25)	3.55 (0.25)
time (week = 12)	$\xi_{1,13}$	4.48 (0.28)	4.42 (0.29)	4.42 (0.29)
antecedents	$\xi_{1,2}$	−0.31 (0.34)	−0.35 (0.34)	−0.35 (0.34)
age/30	$\xi_{1,3}$	−0.18 (0.37)	−0.20 (0.38)	−0.20 (0.38)
duration/100	$\xi_{1,4}$	−0.87 (0.77)	−0.87 (0.78)	−0.85 (0.78)
initial severity	$\xi_{1,5}$	−0.27 (0.21)	−0.25 (0.22)	−0.24 (0.22)
RI sd.	$\sqrt{d_{11}}$	2.35 (0.18)	2.37 (0.19)	1.88 (0.41)
RI sd.	$\sqrt{g_{33}}$	—	—	1.45 (0.50)
Side Effects				
int. 0	$\xi_{2,00}$	−1.71 (1.68)	−2.16 (1.83)	−1.83 (1.84)
int. 1	$\xi_{2,01}$	3.29 (1.69)	2.81 (1.87)	3.17 (1.89)
int. 2	$\xi_{2,02}$	5.64 (1.72)	5.15 (2.01)	5.62 (2.05)
time (week = 4)	$\xi_{2,11}$	0.74 (0.23)	0.66 (0.23)	0.65 (0.23)
time (week = 8)	$\xi_{2,12}$	1.58 (0.26)	1.45 (0.26)	1.45 (0.26)
time (week = 12)	$\xi_{2,13}$	1.64 (0.27)	1.48 (0.27)	1.48 (0.27)
antecedents	$\xi_{2,2}$	−0.12 (0.49)	−0.22 (0.50)	−0.25 (0.51)
age/30	$\xi_{2,3}$	−1.56 (0.54)	−1.58 (0.57)	−1.68 (0.58)
duration/100	$\xi_{2,4}$	−3.97 (1.08)	−3.95 (1.11)	−3.96 (1.12)
initial severity	$\xi_{2,5}$	0.74 (0.30)	0.83 (0.32)	0.79 (0.32)
RI sd.	$\sqrt{g_{22}}$	3.39 (0.30)	3.47 (0.34)	2.29 (0.57)
RI sd.	$\sqrt{g_{44}}$	—	—	2.66 (0.49)
Missing Indicator				
int.	$\xi_{3,0}$	—	−2.90 (0.99)	−3.59 (1.36)
time (week = 4)	$\xi_{3,11}$	—	0.93 (0.36)	1.25 (0.43)
time (week = 8)	$\xi_{3,12}$	—	1.17 (0.37)	1.70 (0.51)
time (week = 12)	$\xi_{3,13}$	—	0.97 (0.40)	1.61 (0.58)
antecedents	$\xi_{3,2}$	—	0.34 (0.29)	0.45 (0.36)
age/30	$\xi_{3,3}$	—	0.39 (0.30)	0.54 (0.38)
duration/100	$\xi_{3,4}$	—	1.67 (0.50)	2.09 (0.69)
initial severity	$\xi_{3,5}$	—	−0.32 (0.16)	−0.41 (0.20)

**Table 4.** Fluvoxamine Trial. Common parameter estimates (s.e.) of the regression coefficients. Model 0: model with two outcomes only; Model 2: model with two random intercepts both of which are shared with missing data indicator; Model 3: model with four random intercepts of which two are shared with missing data indicator

		Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Effect	Par.	Model 0	Model 2	Model 3
Therapeutic Effect & Side Effects				
Cov. RI's	$g_{12}$	0.96 (0.64)	1.17 (0.67)	1.09 (0.67)
Therapeutic Effect & Missing Indicator				
RI sd.	$\sqrt{g_{11}}$	—	2.37 (0.19)	—
RI sd.	$\sqrt{g_{33}}$	—	—	1.45 (0.50)
Scale	$\lambda_1$	—	−1.13 (0.07)	—
Scale	$\lambda_3$	—	—	−0.53 (0.40)
Side Effects & Missing Indicator				
RI sd.	$\sqrt{g_{22}}$	—	3.47 (0.34)	—
RI sd.	$\sqrt{g_{44}}$	—	—	2.66 (0.49)
Scale	$\lambda_2$	—	−0.23 (0.05)	—
Scale	$\lambda_4$	—	—	−0.50 (0.20)

the side effects are included. Further, the missingness process will be associated with the outcomes through these shared random intercepts  $b_{3i}$  and  $b_{4i}$ , which are included using scale factors  $\lambda_3$  and  $\lambda_4$ .

Using the SAS procedure NLMIXED, we fit a joint POMM–POMM–GLM model by relying on numerical integration. **In NLMIXED, not only several integration techniques are available, but also a number of optimization algorithms.** Having experienced problems with the default quasi-Newton optimization algorithm, we replaced it by the more stably converging Newton-Raphson algorithm. Further, given the latter, a numerical sensitivity analysis for different integral approximations was performed. Here, we investigated the influence on model fit not only stemming from the approximation method but also the a priori selected number of quadrature points. More precisely, we initiated the model fitting using *adaptive* Gaussian quadrature approximation for  $Q = 3, 5, 10$ .

The difference between adaptive and non-adaptive Gaussian quadrature is that for the former the quadrature points are appropriately centered and scaled in a way that more quadrature points lie in the region with high probability mass, while for the latter the quadrature points are centered at zero for each of the random effects, and the current random-effects covariance matrix is used as scale matrix (Molenberghs and Verbeke 2005). Clearly, the adaptive Gaussian quadrature approximation is much more time consuming. Subsequently we switched to *non-adaptive* Gaussian quadrature, and fitted the model for  $Q = 10, 20$ . The results are summarized in Tables 5 and 6.

The following observations can be made. Within the adaptive Gaussian quadrature approximation,

there is a minor difference in the parameter estimates (except for some intercepts), also in one of the standard errors, between  $Q = 5$  and  $Q = 10$ . This is also confirmed in terms of log-likelihood. The deviation between the fit for  $Q = 3$  and  $Q = 5, 10$  is much more visible both in terms of the estimates and the log-likelihood value (the  $-2\log$ -likelihood is higher). Switching to non-adaptive approximation and comparing the point and precision estimates and  $-2\log$ -likelihood, we reach practically the same fit for  $Q = 10$  and  $Q = 20$ . These approximations are also numerically close to the adaptive Gaussian quadrature approximation with  $Q = 10$ , and hence, based on the previous comparison, they are also similar to those of the same method but with  $Q = 5$ . However, the most important advantage of replacing the adaptive approximation by the non-adaptive one is the gain in computation time. As all mentioned modifications of model fitting were applied using the same computer platform, a similar way of programming, and the same starting values, we can make a fair comparison of computation times. We observed that almost without losing any quality of the model fit, we can switch from adaptive to non-adaptive Gaussian quadrature, both with  $Q = 10$ . As a result, the computation time decreased from more than 81 hours to approximately 1 hour. However, in the context of the fluvoxamine data and the model considered, adaptive quadrature with  $Q = 5$  is a feasible technique: the quality of the approximation is almost the same as with the latter method and the computation time is a little over 5 hours, which is not unreasonable.

In addition, we investigated whether we can reduce the computation time even more when using as starting parameters for non-adaptive Gaussian approximation with  $Q = 20$  the estimated values from the same procedure but with  $Q = 10$ . However, we observed that this was not a promising route: the model fit was almost the same as for  $Q = 10$  and the implementation ran at approximately the same speed.

Given that the models considered have a likelihood basis, they can be compared using likelihood ratio or related tests when they are nested, or using information criteria when they are not. However, because data are incomplete, one should also investigate the sensitivity of inferences drawn to necessarily non-verifiable assumptions about the predictive distribution, i.e., the distribution of the unobserved data, given observed data, covariates, and the missing data mechanism. Indeed, models with the same or similar fit to the observed data can strongly differ in terms of the predictive distribution (Molenberghs *et al.*, 2008). Verbeke, Molenberghs, and Beunckens (2008) presented formal and informal model selection and model assessment tool for such cases. The fact that we present several models with differing assumptions about the relationships between outcome and missingness processes can be seen as a component of such a sensitivity analysis.

#### 2.4. *Multiple Imputation and Sensitivity Analysis*

Multiple imputation, at least in its basic form (Rubin 1978) requires the missingness mechanism to be MAR. That is why we started our investigation with the ACMV method that works under MAR.

**Table 5.** Fluvoxamine Trial. Response-specific parameter estimates (s.e.) of the regression coefficients of Model 3 with four random intercepts, two of which are shared. The results are listed for adaptive ( $Q=3, 5, 10$ ) and non-adaptive Gaussian quadrature ( $Q=10, 20$ )

Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
		Adaptive			Non-Adaptive	
		$Q = 3$	$Q = 5$	$Q = 10$	$Q = 10$	$Q = 20$
Therapeutic Effect						
int. 0	$\xi_{1,00}$	-1.68 (1.24)	-2.21 (1.23)	-2.27 (1.23)	-2.21 (1.25)	-2.26 (1.23)
int. 1	$\xi_{1,01}$	1.52 (1.28)	0.97 (1.26)	0.91 (1.26)	0.97 (1.28)	0.92 (1.26)
int. 2	$\xi_{1,02}$	3.94 (1.36)	3.37 (1.32)	3.31 (1.32)	3.37 (1.34)	3.31 (1.31)
time (week = 4)	$\xi_{1,11}$	2.05 (0.20)	2.04 (0.20)	2.04 (0.20)	2.05 (0.20)	2.04 (0.20)
time (week = 8)	$\xi_{1,12}$	3.57 (0.25)	3.55 (0.25)	3.55 (0.25)	3.56 (0.25)	3.55 (0.25)
time (week = 12)	$\xi_{1,13}$	4.44 (0.29)	4.42 (0.29)	4.42 (0.29)	4.43 (0.29)	4.42 (0.29)
antecedents	$\xi_{1,2}$	-0.35 (0.34)	-0.35 (0.34)	-0.35 (0.34)	-0.34 (0.34)	-0.35 (0.34)
age/30	$\xi_{1,3}$	-0.28 (0.38)	-0.21 (0.38)	-0.20 (0.38)	-0.20 (0.38)	-0.20 (0.38)
duration/100	$\xi_{1,4}$	-0.86 (0.78)	-0.86 (0.78)	-0.85 (0.78)	-0.86 (0.78)	-0.85 (0.78)
initial severity	$\xi_{1,5}$	-0.34 (0.22)	-0.25 (0.22)	-0.24 (0.22)	-0.25 (0.22)	-0.24 (0.22)
RI sd.	$\sqrt{g_{11}}$	1.93 (0.41)	1.87 (0.40)	1.88 (0.41)	1.93 (0.42)	1.89 (0.41)
RI sd.	$\sqrt{g_{22}}$	1.37 (0.54)	1.46 (0.48)	1.45 (0.50)	1.37 (0.58)	1.43 (0.51)
Side Effects						
int. 0	$\xi_{2,00}$	-1.70 (1.77)	-1.80 (1.82)	-1.83 (1.84)	-0.93 (2.35)	-1.83 (1.83)
int. 1	$\xi_{2,01}$	3.23 (1.82)	3.20 (1.87)	3.17 (1.89)	4.13 (2.48)	3.18 (1.88)
int. 2	$\xi_{2,02}$	5.64 (1.98)	5.63 (2.03)	5.62 (2.05)	6.64 (2.71)	5.62 (2.04)
time (week = 4)	$\xi_{2,11}$	0.66 (0.23)	0.66 (0.23)	0.65 (0.23)	0.66 (0.23)	0.65 (0.23)
time (week = 8)	$\xi_{2,12}$	1.46 (0.26)	1.45 (0.26)	1.45 (0.26)	1.46 (0.26)	1.45 (0.26)
time (week = 12)	$\xi_{2,13}$	1.49 (0.27)	1.48 (0.27)	1.48 (0.27)	1.49 (0.27)	1.48 (0.27)
antecedents	$\xi_{2,2}$	-0.22 (0.49)	-0.24 (0.50)	-0.25 (0.51)	-0.42 (0.52)	-0.25 (0.50)
age/30	$\xi_{2,3}$	-1.64 (0.55)	-1.67 (0.57)	-1.68 (0.58)	-1.91 (0.62)	-1.67 (0.57)
duration/100	$\xi_{2,4}$	-3.91 (1.08)	-3.96 (1.11)	-3.96 (1.12)	-4.02 (0.96)	-3.96 (1.13)
initial severity	$\xi_{2,5}$	0.75 (0.31)	0.78 (0.32)	0.79 (0.32)	0.70 (0.36)	0.77 (0.32)
RI sd.	$\sqrt{g_{33}}$	2.35 (0.46)	2.39 (0.54)	2.29 (0.57)	2.37 (0.45)	2.32 (0.56)
RI sd.	$\sqrt{g_{44}}$	2.40 (0.48)	2.50 (0.50)	2.66 (0.49)	2.67 (0.40)	2.64 (0.50)
Missing Indicator						
int.	$\xi_{3,0}$	-3.67 (1.37)	-3.61 (1.37)	-3.59 (1.36)	-3.82 (1.41)	-3.61 (1.37)
time (week = 4)	$\xi_{3,11}$	1.24 (0.43)	1.25 (0.43)	1.25 (0.43)	1.25 (0.43)	1.25 (0.44)
time (week = 8)	$\xi_{3,12}$	1.69 (0.51)	1.70 (0.51)	1.70 (0.51)	1.70 (0.50)	1.71 (0.51)
time (week = 12)	$\xi_{3,13}$	1.60 (0.58)	1.62 (0.58)	1.61 (0.58)	1.62 (0.57)	1.62 (0.58)
antecedents	$\xi_{3,2}$	0.45 (0.37)	0.45 (0.37)	0.45 (0.36)	0.49 (0.37)	0.45 (0.37)
age/30	$\xi_{3,3}$	0.55 (0.38)	0.54 (0.38)	0.54 (0.38)	0.59 (0.38)	0.54 (0.38)
duration/100	$\xi_{3,4}$	2.10 (0.70)	2.10(0.70)	2.09 (0.69)	2.10 (0.67)	2.10 (0.69)
initial severity	$\xi_{3,5}$	-0.39 (0.20)	-0.41 (0.20)	-0.41 (0.20)	-0.38 (0.21)	-0.41 (0.20)

**Table 6.** Fluvoxamine Trial. Common parameter estimates (s.e.) of the regression coefficients of Model 3 with four random intercepts, two of which are shared. The results and computation time are listed for adaptive ( $Q=3, 5, 10$ ) and non-adaptive Gaussian quadrature ( $Q=10, 20$ )

		Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
		Adaptive			Non-Adaptive	
Effect	Par.	$Q = 3$	$Q = 5$	$Q = 10$	$Q = 10$	$Q = 20$
Therapeutic Effect & Side Effects						
Cov. RI's	$g_{12}$	1.05 (0.64)	1.06 (0.66)	1.09 (0.67)	0.88 (0.71)	1.07 (0.68)
Therapeutic Effect & Missing Indicator						
RI sd.	$\sqrt{g_{22}}$	1.37 (0.54)	1.46 (0.48)	1.45 (0.50)	1.37 (0.58)	1.43 (0.51)
Scale	$\lambda_1$	-0.58 (0.45)	-0.49 (0.35)	-0.53 (0.40)	-0.56 (0.49)	-0.54 (0.41)
Side Effects & Missing Indicator						
RI sd.	$\sqrt{g_{44}}$	2.40 (0.48)	2.50 (0.50)	2.66 (0.49)	2.67 (0.40)	2.64 (0.50)
Scale	$\lambda_4$	-0.58 (0.22)	-0.57 (0.23)	-0.50 (0.20)	-0.51 (0.17)	-0.52 (0.20)
-2ll		3981.0	3975.4	3974.3	3973.7	3974.2
CPU		42:07	5:07:30	81:26:54	1:00:04	15:23:35

Given that current computational facilities support larger numbers of  $M$  to be used at minimal cost (except maybe for very large datasets), it seems reasonable to choose  $M$  in such a way that the final analysis is reproducible, i.e., if one wants to repeat the analysis for the same  $M$  then essentially the same results will be generated. Hence, all MI methods implemented by us were ran for  $M = 20$  imputations. In the Supplementary Materials A, we show that  $M = 10$  returns essentially the same results.

The ACMV analysis was implemented as follows: first the procedure MI (SAS 9.4) is used for the imputations, then every completed data is analysed using procedure NLMIXED for the joint modelling of the therapeutic and side effects with the general assumption for the covariance matrix (2), and finally the results of the separate analysis are combined into a single one using the MIANALYZE procedure.

Further, one might want to examine how the MAR and MNAR models differ in their prediction of unobserved data. Another interesting point is how the results depend on the way the multivariate distribution of the longitudinal process is treated. To compare the results of the analysis, given the different hypotheses about the identifying restriction for missing data, we will perform a sensitivity analysis for the multiple imputation (Carpenter *et al.* 2013).

Since at the end of the study, as a number of patients dropped out, the variability of side effects and therapeutic effect turned out to be insufficient to include both variables into the MI analysis. Given the fact that with high side effects a reasonable therapeutic effect is absent, we decided to remove the last (*i.e.*, the 5th) measurement of the therapeutic effect in the MI analysis.

Based on the joint density for the observed and missing data, we will perform the MI analysis given identifying restrictions MI 1–3. MI 4 is an identifying restriction based on a conditional distribution  $f(\mathbf{y}_j | \mathbf{y}_1, \dots, \mathbf{y}_{j-1})$ , where for each variable  $\mathbf{Y}_j$  we assume it to be the same for  $\mathbf{Y}_j^o$  and  $\mathbf{Y}_j^m$ :

- (MI 1) under MAR: *available cases missing value* (ACMV) method.
- (MI 2) under MNAR: *neighboring case missing values* (NCMV). In this method, for the imputed missing values of variable  $\mathbf{Y}_j$ , we estimated the posterior density based on the patterns for which  $\mathbf{Y}_j$  is observed and  $\mathbf{Y}_{j+1}$  is missing. The mentioned principle for identifying  $\mathbf{Y}_j$  can be extended to larger numbers of patterns: the patterns for which  $\mathbf{Y}_j$  is observed and  $\mathbf{Y}_{j+k}$  is missing.
- (MI 3) under MNAR: *complete case missing value* (CCMV).
- (MI 4) under MAR: the *Full Conditional Specification* (FCS; van Buuren 2007) method was applied. Here, it was assumed that, if the patient left the study, all his/her “after dropout” measurements are equal to the overall mean.

All aforementioned modifications of the MI analysis (MI 1–4) were implemented in SAS (SAS 9.4) in the same way as described in the beginning of this section. The results for the point and precision estimates are listed in Table 7. We note that FCS and ACMV are closest to each other in terms of estimates and precision estimates. This is to be expected, given their MAR basis. CCMV generates approximately the same results except for the category-specific intercepts: they are slightly different but have almost the same precision as with FCS and ACMV.

When compared with CCMV, FCS, and ACMV, NCMV generates slightly different approximations for the fixed effects and also for the standard deviations of the random intercepts. However, the largest difference lies in the estimation of the covariance between the random intercepts of the two responses: for NCMV the covariance is much higher. It can be explained by NCMV: here, we defined the posterior density by the neighboring patterns, which by nature are more similar to each other, hence, the higher correlation between the joint responses. The precision estimates of all parameters for NCMV are approximately the same as for the other methods.

### 3. Discussion

In this article, we proposed an application of the joint modelling of composite endpoint method on the fluvoxamine trial, a study with two longitudinal ordinal outcomes where not all patients remained in the study. Since the therapeutic and side effects were measured simultaneously, they were regarded as components of a composite endpoint and had to be modelled jointly. Then, we extended our model by incorporating missingness in various ways to estimate the effect on the results.

After applying several MNAR models, we concluded that they yield approximately the same fit as their MAR counterpart but with a slightly higher uncertainty. This ensures that the corresponding MAR version properly fits the observed data. For simplicity, the models formulated include only

**Table 7.** Fluvoxamine Trial. Parameter estimates (s.e.) of the regression coefficients for available case missing value (ACMV), neighboring case missing value (NCMV), complete case missing value (CCMV) and full conditional specification (FCS) multiple imputation. All are performed for  $M = 20$  imputations

		Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Effect	Par.	ACMV	NCMV	CCMV	FCS
Therapeutic Effect					
int. 0	$\xi_{1,00}$	-2.03 (1.14)	-2.27 (1.13)	-1.96 (1.10)	-2.00 (1.08)
int. 1	$\xi_{1,01}$	1.01 (1.15)	0.56 (1.13)	1.13 (1.10)	1.06 (1.08)
int. 2	$\xi_{1,02}$	3.40 (1.16)	2.75 (1.13)	3.55 (1.10)	3.45 (1.10)
time (week = 4)	$\xi_{1,11}$	1.99 (0.19)	1.74 (0.18)	2.06 (0.19)	2.01 (0.19)
time (week = 8)	$\xi_{1,12}$	3.50 (0.24)	2.60 (0.21)	3.62 (0.24)	3.52 (0.24)
time (week = 12)	$\xi_{1,13}$	4.32 (0.27)	3.86 (0.25)	4.41 (0.27)	4.34 (0.27)
antecedents	$\xi_{1,2}$	-0.27 (0.31)	-0.65 (0.32)	-0.24 (0.32)	-0.24 (0.31)
age/30	$\xi_{1,3}$	-0.13 (0.36)	-0.37 (0.35)	-0.03 (0.35)	-0.11 (0.34)
duration/100	$\xi_{1,4}$	-0.75 (0.72)	-1.13 (0.74)	-0.64 (0.72)	-0.73 (0.71)
initial severity	$\xi_{1,5}$	-0.29 (0.20)	-0.09 (0.20)	-0.35 (0.19)	-0.31 (0.19)
RI sd.	$\sqrt{g_{11}}$	2.24 (0.17)	2.34 (0.17)	2.25 (0.16)	2.24 (0.17)
Side Effects					
int. 0	$\xi_{2,00}$	-0.72 (1.55)	-1.25 (1.57)	-0.44 (1.48)	-0.94 (1.57)
int. 1	$\xi_{2,01}$	3.87 (1.53)	2.78 (1.59)	4.29 (1.49)	3.72 (1.57)
int. 2	$\xi_{2,02}$	6.54 (1.55)	4.90 (1.63)	6.84 (1.53)	6.48 (1.59)
time (week = 4)	$\xi_{2,11}$	0.76 (0.23)	0.42 (0.21)	0.91 (0.22)	0.78 (0.22)
time (week = 8)	$\xi_{2,12}$	1.51 (0.24)	0.59 (0.23)	1.69 (0.25)	1.52 (0.24)
time (week = 12)	$\xi_{2,13}$	1.59 (0.26)	0.80 (0.23)	1.71 (0.25)	1.62 (0.25)
antecedents	$\xi_{2,2}$	-0.03 (0.45)	-0.52 (0.46)	0.07 (0.44)	0.01 (0.46)
age/30	$\xi_{2,3}$	-1.49 (0.49)	-1.37 (0.50)	-1.47 (0.47)	-1.50 (0.49)
duration/100	$\xi_{2,4}$	-3.81 (0.99)	-4.74 (1.02)	-3.48 (0.95)	-4.00 (1.00)
initial severity	$\xi_{2,5}$	0.52 (0.28)	0.71 (0.27)	0.44 (0.26)	0.56 (0.27)
RI sd.	$\sqrt{g_{22}}$	3.10 (0.30)	3.26 (0.29)	3.08 (0.26)	3.19 (0.28)
Common parameter					
Therapeutic Effect & Side Effects					
Cov. RI's	$g_{12}$	0.83 (0.53)	2.58 (0.66)	0.86 (0.53)	0.90 (0.57)



random intercepts. They can be extended to models with more complex random effect structures, e.g., models with random intercepts and random slopes. Potentially, this may increase complexity of convergence as well as computation time. We paid specific attention to the complexity of the POMM–POMM–GLM MNAR model. We found that an acceptable model fit and reasonable computation time can be obtained for non-adaptive Gaussian quadrature numerical approximation with  $Q = 10$ . Also, adaptive quadrature with  $Q = 5$  is feasible in the context of the study. While adaptive quadrature with a sufficiently large number of quadrature points is recommended, the achieved accuracy and numerical feasibility will depend on the application at hand.

Given the assumptions behind the posterior distribution of the dropouts, different modifications of MI analysis were introduced. As such assumptions are partly untestable, conducting a sensitivity analysis is primordial. Starting with the basic MI analysis in its MAR form (Rubin 1978) and then extended to MNAR, under a suite of the identifying restrictions (Verbeke and Molenberghs 2000, Molenberghs and Kenward 2007, van Buuren 2007), we concluded that, based on the performed sensitivity analysis, in the context of the fluvoxamine trial, inferences regarding some parameters can be modified by some identifying restrictions. After comparing all MI analysis performed, we conclude that the largest difference in inference is for NCMV and is due to the much larger estimates of the covariance between the random intercepts of the two responses. This can be explained by the definition of the posterior density as it considers neighboring patterns which are bound to be similar.

The guideline for choosing the number of imputations is to render the final analysis reproducible, which is not always the case for a small number of imputations (White *et al.* 2011, van Buuren 2012). Hence, a sufficiently large number of imputations is advised (e.g.,  $M = 20$ ) and supported by the currently available computational power.

## Acknowledgments

Financial support from the IAP research network #P7/06 of the Belgian Government (Belgian Science Policy), the Flemish Supercomputer Project, and the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT), in which Intel and Janssen Pharmaceutica are partners, is gratefully acknowledged. We are also grateful to Dr. Kris Bogaerts of I-BioStat for his expert advice.

## References

- Agresti, A., Lang J.B. (1993) A proportional odds model with subject-specific effects for repeated ordered categorical responses. *Biometrika*, **80**, 527–534.
- Breslow, N.E., and Clayton, D.G. (1993) Approximate inference in generalized linear mixed models. *J. Am. Statist. Ass.*, **88**, 9–25.

- Carpenter, J.R. and Kenward, M.G. (2013) *Multiple Imputation and Its Application*. Chichester: John Wiley & Sons.
- Carpenter, J.R., Roger, J.H. and Kenward, M.G. (2013) Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. *J. Biopharm. Statist.*, **23:6**, 1352–1371.
- Creemers, A, Hens, N., Aerts, M., Molenberghs, G., Verbeke, G. and Kenward, M.G. (2010) A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Niometrical Journal* **52** (1), 111–125.
- Engel, B. and Keen, A. (1994) A simple approach for the analysis of generalized linear mixed models. *Statist. Neerl.*, **48**, 1–22.
- Fieuws, S. and Verbeke, G. (2006) Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*, **62**, 424–431.
- Fieuws, S., Verbeke, G., Boen, F. and Delecluse, C. (2006) High-dimensional multivariate mixed models for binary questionnaire data. *Appl. Statist.*, **55**, 1–12.
- Fieuws, S., Verbeke, G., Maes B., and Vanrenterghem Y. (2008) Predicting renal graft failure using multivariate longitudinal profiles. *Biostatistics*, **9**, 419–431.
- Fitzmaurice, G., Laird, N., and Ware, J. (2004) *Applied Longitudinal Analysis*. Hoboken: Wiley.
- Follmann, D. and Wu, M. (1995) An approximate generalized linear model with random effects for informative missing data. *Biometrics*, **51**, 151–68.
- Horrocks, J. and van den Heuvel, M.J. (2009) Prediction of pregnancy: A joint model for longitudinal and binary data. *Baysn Anal.*, **4**, 523–538.
- Iddi, S. and Molenberghs, G. (2012) A joint marginalized multilevel model for continuous and binary longitudinal outcomes. *J. Appl. Statist.*, **00**, 000–000.
- Kenward, M.G., Lesaffre, E. and Molenberghs, G. (1994) An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics*, **50**, 945–953.
- Laird, N. and Ware, J. (1982) Random-effects models for longitudinal data. *Biometrics*, **38**, 963–974.
- Little, R.J.A. (1993). Pattern-mixture models for multivariate incomplete data. *J. Am. Statist. Ass.*, **88**, 125–34.
- Little, R.J.A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika*, **81**, 471–483.

- Little, R.J.A. (1995) Modelling the drop-out mechanism in repeated-measures studies. *J. Am. Statist. Ass.*, **90**, 1112–21.
- Little, R.J.A., and Rubin, D.B. (2002) *Statistical Analysis with Missing Data*. 2nd ed. Chichester, UK: Wiley.
- Michiels, B. and Molenberghs, G. (1997) Protective estimation of longitudinal categorical data with nonrandom dropout. *Commun Statist - Theor Method*, **26**, 65–94.
- Molenberghs, G., Beunckens, C., Sotito, C. and Kenward, M.G. (2008) Every missing not at random model has got a missing at random counterpart with equal fit. *J. R. Statist. Soc. B*, **70**, 371–388.
- Molenberghs, G. and Kenward, M.G. (2007) *Missing Data in Clinical Studies*. New York: Wiley.
- Molenberghs, G., Kenward, M.G., and Lesaffre, E. (1997) The analysis of longitudinal ordinal data with non-random dropout. *Biometrika*, **84**, 33–44.
- Molenberghs, G., and Lesaffre, E. (1994) Marginal modelling of correlated ordinal data using a multivariate Plackett distribution. *J. Am. Statist. Ass.*, **89**, 633–644.
- Molenberghs, G. and Verbeke, G. (2005) *Models for Discrete Longitudinal Data*. New York: Springer.
- Njagi, E.N., Molenberghs, G., Kenward, M.G., Verbeke, G. and Rizopoulos, D. (2014) A characterization of missingness at random in a generalized shared-parameter joint modeling framework for longitudinal and time-to-event data, and sensitivity analysis. *Biom. J.*, **56**(6), 1001–1015.
- Rizopoulos, D. (2012) *Joint Models for Longitudinal and Time-to-Event Data with Applications in R*. Boca Raton: Chapman & Hall/CRC.
- Rubin, D.B. (1976) Inference and missing data. *Biometrika*, **63**, 581–92.
- Rubin, D.B. (1978) Multiple imputations in sample surveys – a phenomenological Bayesian approach to nonresponse. In: *Imputation and Editing of Faulty or Missing Survey Data*. Washington, DC: U.S. Department of Commerce, pp.1–23. *Biometrika*, **63**, 581–92.
- Rubin, D.B. (1987) *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Rubin, D.B. and Schenker, N. (1986) Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *J. Am. Statist. Ass.*, **81**, 366–374.
- Schafer, J. (2003) Multiple imputation in multivariate problems when the imputation and analysis models differ. *Statist. Neerl.*, **57**, 19–35.
- Thiébaud, R., Jacqmin-Gadda, H., Chêne, G., Lepout, C. and Commenges, D. (2002). Bivariate linear mixed models using SAS PROC MIXED. *Comput Meth Prog Bio*, **69**, 249–256.

- Tsiatis, A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statist. Sinica*, **14**, 809–834.
- van Buuren, S. (2007) Multiple imputation of discrete and continuous data by fully conditional specification. *Statist. Meth. Med. Res.*, **16**, 219–242.
- van Buuren, S. (2012) *Flexible Imputation of Missing Data*. Boca Raton: CRC/Chapman & Hall.
- Vasdekis, V.G.S, Rizopoulos, D. and Moustaki I.(2014) A weighted pairwise likelihood estimator for a class of latent variable models. *Biostatistics* **15** (4), 677–689.
- Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data*. New York: Springer.
- Verbeke, G., Molenberghs, G., and Beunckens, C. (2008) Formal and informal model selection with incomplete data. *Statistical Science*, **23**, 201–218.
- White, I.R., Royston, P. and Wood, A.M. (2011) Multiple imputation using chained equations: issues and guidance for practice. *Statist. Med.*, **30**, 377–399.
- Wolfinger, R. and O’Connell, M. (1993) Generalized linear mixed models: a pseudo-likelihood approach. *J. Statist. Computn. Simuln.*, **48**, 233–243.
- Wu, M.C., and Carroll, R.J. (1988) Estimation and comparison of changes in the presence of informative right censoring by modelling the censoring process. *Biometrics*, **44**, 175–88.

# **Mechanism for Missing Data Incorporated into Joint Modelling of Ordinal Responses**

Anna Ivanova

*I-BioStat, KU Leuven, Belgium*

Geert Molenberghs

*I-BioStat, Universiteit Hasselt, Belgium; I-BioStat, KU Leuven, Belgium*

Geert Verbeke

*I-BioStat, KU Leuven, Belgium; I-BioStat, Universiteit Hasselt, Belgium*

## **Supplementary Materials**

### **A. Additional Table**

**Table 8.** Fluvoxamine Trial. Parameter estimates (s.e.) of the regression coefficients for available case missing value (ACMV), neighboring case missing value (NCMV), complete case missing value (CCMV) and full conditional specification (FCS) multiple imputation. All are performed for  $M = 10$  imputations

		Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Effect	Par.	ACMV	NCMV	CCMV	FCS
Therapeutic Effect					
int. 0	$\xi_{1,00}$	-1.98 (1.10)	-2.29 (1.12)	-2.09 (1.08)	-2.04 (1.11)
int. 1	$\xi_{1,01}$	1.06 (1.11)	0.54 (1.12)	0.99 (1.07)	1.04 (1.12)
int. 2	$\xi_{1,02}$	3.44 (1.12)	2.69 (1.13)	3.39 (1.09)	3.45 (1.13)
time (week = 4)	$\xi_{1,11}$	2.00 (0.19)	1.72 (0.18)	2.07 (0.19)	2.02 (0.19)
time (week = 8)	$\xi_{1,12}$	3.48 (0.25)	2.60 (0.20)	3.62 (0.23)	3.53 (0.24)
time (week = 12)	$\xi_{1,13}$	4.32 (0.27)	3.84 (0.24)	4.40 (0.27)	4.36 (0.27)
antecedents	$\xi_{1,2}$	-0.28 (0.31)	-0.62 (0.32)	-0.21 (0.31)	-0.24 (0.32)
age/30	$\xi_{1,3}$	-0.12 (0.35)	-0.39 (0.35)	-0.01 (0.34)	-0.11 (0.35)
duration/100	$\xi_{1,4}$	-0.69 (0.72)	-1.29 (0.72)	-0.83 (0.70)	-0.85 (0.72)
initial severity	$\xi_{1,5}$	-0.30 (0.19)	-0.07 (0.19)	-0.32 (0.19)	-0.30 (0.20)
RI sd.	$\sqrt{g_{11}}$	2.23 (0.17)	2.32 (0.17)	2.23 (0.17)	2.29 (0.17)
Side Effects					
int. 0	$\xi_{2,00}$	-0.85 (1.51)	-1.17 (1.53)	-0.50 (1.46)	-1.02 (1.57)
int. 1	$\xi_{2,01}$	3.75 (1.50)	2.90 (1.54)	4.23 (1.48)	3.68 (1.58)
int. 2	$\xi_{2,02}$	6.42 (1.54)	5.04 (1.55)	6.77 (1.52)	6.35 (1.64)
time (week = 4)	$\xi_{2,11}$	0.75 (0.23)	0.43 (0.20)	0.89 (0.21)	0.76 (0.21)
time (week = 8)	$\xi_{2,12}$	1.50 (0.26)	0.62 (0.22)	1.69 (0.24)	1.52 (0.24)
time (week = 12)	$\xi_{2,13}$	1.61 (0.26)	0.84 (0.22)	1.77 (0.23)	1.61 (0.25)
antecedents	$\xi_{2,2}$	-0.03 (0.44)	-0.52 (0.48)	0.11 (0.45)	0.06 (0.44)
age/30	$\xi_{2,3}$	-1.49 (0.50)	-1.45 (0.50)	-1.48 (0.48)	-1.52 (0.48)
duration/100	$\xi_{2,4}$	-3.87 (1.01)	-4.81 (1.03)	-3.60 (0.96)	-4.09 (0.99)
initial severity	$\xi_{2,5}$	0.55 (0.27)	0.72 (0.27)	0.46 (0.26)	0.58 (0.27)
RI sd.	$\sqrt{g_{22}}$	3.12 (0.30)	3.28 (0.26)	3.11 (0.28)	3.19 (0.26)
Common parameter					
Therapeutic Effect & Side Effects					
Cov. RI's	$g_{12}$	0.84 (0.53)	2.57 (0.63)	0.81 (0.54)	1.02 (0.60)

## B. Variance-Covariance Matrices of Random-Effects Structures

**Model 1:** The  $b_{3i}$  are pairwise correlated with  $b_{1i}$  and  $b_{2i}$ , leading to:

$$G = \begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{12} & g_{22} & g_{23} \\ g_{13} & g_{23} & g_{33} \end{pmatrix} \quad (4)$$

**Model 2:** The random intercepts for both outcomes,  $b_{1i}$  and  $b_{2i}$ , are shared with the missing-data indicator, and scaled by factors  $\lambda_1$  and  $\lambda_2$ .

$$G = \begin{pmatrix} g_{11} & g_{12} & \lambda_1 g_{11} \\ g_{12} & g_{22} & \lambda_2 g_{22} \\ \lambda_1 g_{11} & \lambda_2 g_{22} & \lambda_1^2 g_{11} + \lambda_2^2 g_{22} + 2\lambda_1 \lambda_2 g_{12} \end{pmatrix} \quad (5)$$

**Model 3:** The random intercepts are  $b_{3i}$  for therapeutic effect and  $b_{4i}$  for side effects; missingness is associated with the outcomes through  $b_{3i}$  and  $b_{4i}$  and included with scale factors  $\lambda_3$  and  $\lambda_4$ , respectively.

$$G = \begin{pmatrix} g_{11} + g_{33} + 2g_{13} & g_{12} + g_{14} + g_{23} + g_{24} & \lambda_1 g_{13} + \lambda_1 g_{33} + \lambda_2 g_{14} + \lambda_2 g_{34} \\ g_{12} + g_{14} + g_{23} + g_{24} & g_{22} + g_{44} + 2g_{24} & \lambda_1 g_{23} + \lambda_1 g_{34} + \lambda_2 g_{24} + \lambda_2 g_{44} \\ \lambda_1 g_{13} + \lambda_1 g_{33} + \lambda_2 g_{14} + \lambda_2 g_{34} & \lambda_1 g_{23} + \lambda_1 g_{34} + \lambda_2 g_{24} + \lambda_2 g_{44} & \lambda_1^2 g_{33} + \lambda_2^2 g_{34} + 2\lambda_1 \lambda_2 g_{34} \end{pmatrix} \quad (6)$$